REPORT DOCUMENTATION PAGE

AD-A272 683

2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
05 Nov 93	Final 01 Apr 93 - 30 Sep 93

4. TITLE AND SUBTITLE

An Optically-Assisted 3-D Cellular Array Machine

5. FUNDING NUMBERS

C: N00014-93-C-0021

6. AUTHOR(S)

Freddie Lin

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Physical Optics Corporation
20600 Gramercy Place, Building 100
Torrance, California 90501

8. PERFORMING ORGANIZATION REPORT NUMBER

3210

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Office of Naval Research
Code 1512B : JGW
Ballston Tower One
800 North Quincy Street, Arlington, VA 22217-5660

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

DTIC
ELECTE
NOV 16 1993
S
A

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release
Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

In order to increase the real-time performance of next-generation image processing systems, Physical Optics Corporation (POC) explored a new architectural structure -- a hybrid optically-assisted 3-D cellular array machine. Three techniques are used in this architecture. An analog processing technique is used for front-end high-speed image preprocessing. A digital processing technique is employed for high-precision and highly programmable image processing. An optical interconnect technique is used for high-efficiency high-bandwidth data communication between the analog and digital processing layers. The combination of these three techniques yields a real-time image processing system in a compact 3-D system package with scalable expansion capability.

*Original contains color
plates: All DTIC reproduct-
ions will be in black and
white*

14. SUBJECT TERMS
Image Processing, Cellular Array Machine, Optical Interconnects, Digital Processing, Analog Processing

15. NUMBER OF PAGES
37

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	SAR

AN OPTICALLY-ASSISTED 3-D CELLULAR ARRAY MACHINE

Final Report

Contract No. N00014-93-C-0021
Period of Performance: April 1, 1993 to September 30, 1993

Presented to:
Office of Naval Research
Ballston Tower One
800 North Quincy Street
Arlington, Virginia 22217-5660

Scientific Officer:
Dr. Clifford Lau
Code 1114SE

Presented by:
Physical Optics Corporation
Research & Development Division
20600 Gramercy Place, Suite 103
Torrance, California 90501

Principal Investigator
Freddie Lin, Ph. D.
(310) 320-3088

September 1993

TABLE OF CONTENTS

i

1.0 INTRODUCTION

In past decades, significant progress has been made in understanding the first steps in visual processing. Thus, a large number of well studied algorithms exist to locate edges, compute disparities along these edges or over areas, estimate motion fields and find discontinuities in depth, motion, color and texture. Many of these algorithms are formulated as relaxation algorithms which need to be executed many hundreds of times before convergence. Applying these algorithms to all the pixels of an image in real-time is quite a challenge for the modern computer. Real-time image processing requires enormously high data throughput. Concurrent or parallel processing methods appear to be the only means to achieve these processing throughputs. Additional constraints, if these systems are to find widespread use, are that they must be moderate in cost and size. These constraints preclude the use of general purpose supercomputers such as the Cray XMP and the Hitachi S-810.

A cellular array "machine" approach, using a nearest neighbor interconnect type, has been used to solve the problem of real-time image processing with high data throughput. The cellular array machine assigns an individual processor to each image pixel, or pixel subarray, of the entire input image data with only local interconnects allowed between processors. These locally interconnected architectures can perform real-time processing with no degradation in the frame rate. To achieve a full-scale system using this approach, a number of design problems need to be addressed:

1. Processor Element Design. The design of the processor should include its central processor, memory, and I/O ports. The processor design *should be as flexible as possible* in order to handle a wide variety of image processing algorithms. Most importantly, the communication throughput between processor elements should be very high to meet the real-time processing requirement of the system.

2. Architectural Design. In particular, the data communication between a host computer and the cellular array machine, and within the cellular array machine, needs to be investigated. Different electronic processing techniques (digital and analog) should also be explored in the same architectural design.

3. Additional Support. Additional support is required to enhance the performance of the system. This will allow the scaleable expansion of the system.

93-27939

1

93 11 12 102

To address these design problems, Physical Optics Corporation (POC), which specializes in signal processing techniques using both photonic and electronic technologies, proposed to develop an optically-assisted 3-D cellular array machine, associated with analog and digital VLSI chips, and opto-electronic chips, for the development of a full scale, real-time image processing system. The proposed chip set utilizes the high-data bandwidth of optical interconnect components while maintaining the processing flexibility of VLSI electronic processing circuits.

Figure 1-1 is the schematic diagram of the proposed 3-D cellular array machine. An electronic layer consists of an array of modularized cellular processing nodes. Each processing module has three units: an electronic processing element for image data processing, conditioning and memory; a light source circuit for delivering processed signals from one layer to an adjacent layer; and a light detection circuit for receiving data from an adjacent board. The connection between layers can be either free-space optical interconnects, i.e., electronic signals converted to light signals and transmitted in free space to an adjacent layer at the corresponding (x,y) position, or it can be transmitted by optical fibers to another processor in a remote location.
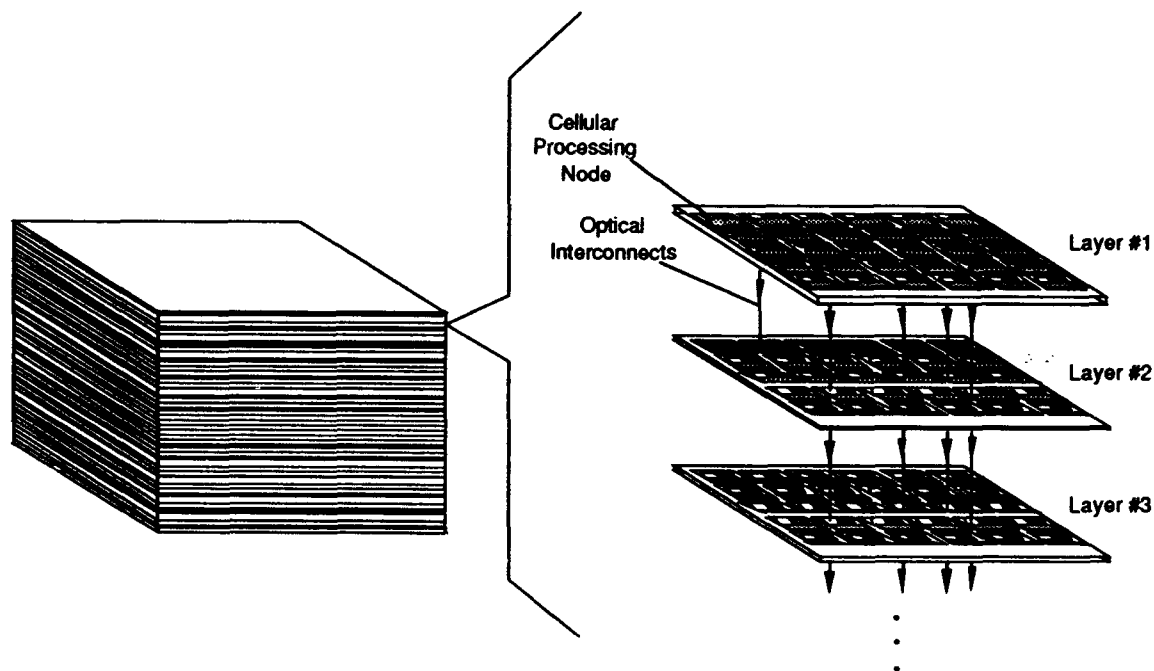
Figure 1-1
Overview of the proposed optically-assisted 3-D cellular array machine.

Both digital and analog electronic techniques can be used to implement the cellular array machine. The analog technique is suitable for fast, low-level image pre processing, while the digital technique can be used for high-level and highly programmable image processing. The best solution is a system combining both analog and digital processing. The analog elements will be used in the front-end unit (layer #1 in Figure 1-1), performing high speed preprocessing. Then, the digital processing elements resident in the other layers of the 3-D cellular array machine will focus on the critical areas of the image, executing high precision algorithms.

During our Phase I program, we designed a hybrid (digital and analog) image processing system which will incorporate an optical interconnect technique to realize high-data bandwidth communication channels. We conclude that a very substantial gain in processing performance can be achieved through the combination of analog preprocessing units, digital processing units and opto-electronic interconnect units. In this final report for the Phase I project, we discuss the details of the system architecture, the building block components: analog units, digital processors, and optical links.

2.0   ARCHITECTURE

There are currently many techniques used in image processing. The conventional approach is a digital image processor, generally a serial processor. In this technique, the image is sensed, quantized and processed pixel by pixel. The advantage of digital processing is that it can adopt some well known algorithms and will have high precision (256 levels or higher). However, a longer time is required to attain this precision. Although some parallel image processing systems have been reported [1], they suffer from several limitations, such as the large number of processors and control logic units needed to achieve real-time operation.

The other approach is to use analog processors. The computing architecture of this implementation consists of massively parallel interconnections of simple neural processors [2]. The inherent parallelism yields fast operation. Moreover, this architecture consumes less power and occupies a smaller area due to its simpler structure. One of the well-known types is the "silicon retina" proposed by Carver Mead [3].

More recently, a new technique called the cellular neural network (CNN) was proposed by Chua and Yang [4]. This is a special type of analog nonlinear processor array which is comprised of a two-dimensional array of identical, equally spaced, processing elements which are interconnected

3

directly to their nearest neighbors. The local connectivity simplifies the layout. In addition, it also limits the number of inputs from other cells. This is the problem for most VLSI implementation of analog neural networks for which saturation and cumulative inaccuracy are often problems. Analog CNN circuits are very effective in real-time image processing applications such as noise removal, edge detection, and feature extraction. The local connectivity makes it suitable for VLSI implementation [5].

The potential disadvantage of both of the analog processors is the lower computation precision due to the inherent properties of the MOS transistor. Pixels with higher or lower intensity tend to saturate at the output of the MOS transistor. This makes the analog processor unsuitable for images with low contrast or images in a highly cluttered background. However, because they are simple and fast, the analog processors are still good candidates for image preprocessing. Any incomplete portions in the pre-processed image can be further processed by digital processors. Combining the two types of processors can achieve both fast and accurate results.

After studying the literature on visual image processing system architectures, we came to the conclusions that:

* Analog VLSI imaging processing techniques, such as silicon retinas, CNN, or early-vision neural chips, are most suitable for front-end early-vision processing because of their advantages in processing throughput, chip area, and power consumption.

* The two principal drawbacks of analog VLSI imaging techniques are their lack of programming flexibility and their inaccuracy. These techniques are usually hardwired to perform specific work and are difficult to program in real time. Some techniques are designed to be programmable (e.g., the Universal CNN [6-14]). These techniques will significantly reduce the processing burden of the digital imaging processors. However, the accuracy of these analog techniques is low. Thus, they may not be suitable for processing low-contrast, high-clutter images.

* The incorporation of both digital and analog VLSI image processing techniques produces a system which combines the advantages of both techniques.

Based on these conclusions, we initiated an architectural design for a real-time image processing system. The system consists of:

1. Analog processors to perform fast, low-level image processing.

2. Multi-layered digital processors to provide high-precision, high-level processing.

3. High bandwidth optical interconnects to achieve efficient communication between processing layers.

Figure 2-1 shows a block diagram of the architectural design produced in Phase I. The system consists of a layer of analog VLSI processing layers and a number of digital VLSI processing layers. These layers are stacked together to form a 3-D configuration. Communication between layers is realized by optical interconnects such that the high data transmission throughput can be maintained at the layer-to-layer level. The key function of the analog VLSI processing layer is to perform high-speed, but coarse, early, visual information processing. The processed image data from the analog VLSI processing layer provides information to the digital VLSI processing layer for selective or prioritized fine image processing. In other words, the analog VLSI layer performs quick coarse image processing operations on the received image and directs the digital VLSI layer to the critical regions of the received image. These critical regions are the low-contrast high-clutter areas which may be overlooked by the analog processing layer. In this way, the digital VLSI layer does not have to process every pixel in the received image and can focus its processing power, flexible programming, and accurate calculations only on the critical regions.

The analog VLSI processing layer consists of an array of analog VLSI image processing nodes interconnected via a cellular array. Each node contains three parts: a photodetector array for object imaging, an analog VLSI processing circuit for high-speed, early visual information processing, and photonic interface unit. The designs of the photodetector array and analog VLSI processing circuit are similar to cellular neural networks, silicon retinas, or early-vision neural chips. The function of photonic interface unit is to convert electronic signals to optical signals for high-throughput, parallel, layer-to-layer communication. Note that the analog VLSI layer transmits both the processed image data (as the control data for the digital VLSI layers) and the unprocessed raw image data (as the image data for the digital VLSI layers) to the digital VLSI layers.
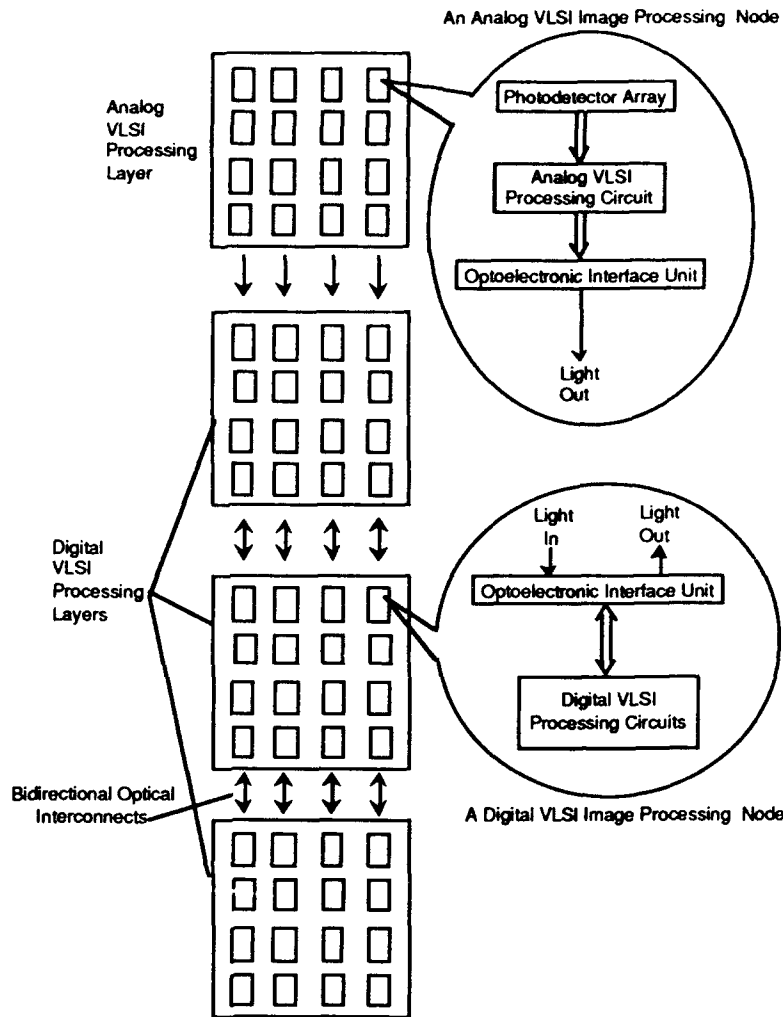
Figure 2-1
Image processing architecture which combines analog with digital technologies. The analog layer
is used for fast but coarse preprocessing, the digital layers for high precision and high level
processing.

The digital VLSI layers are arranged in a multi-layer cellular array configuration with digital VLSI
image processing nodes interconnected by electronic wires within the layer and by optical
interconnects between layers. Each digital VLSI image processing node consists of a photonic
interface unit for bi-directional communication and a digital VLSI processing circuit for flexible,
high-precision image processing.

A host computer will handle the control and other operations. The host computer will include a hard disk, a floppy disk drive, one or more video displays, a keyboard for operator interaction, and a VME-bus or PC-bus backplane for peripheral interfaces.

2.1 Analog Front-End Unit

During the Phase I project, a number of analog image processing techniques were studied. The cellular neural network (CNN) [6-14] was chosen for further study and Phase II implementation. The CNN has three unique features: simple silicon fabrication process, large processing power and programmability.

The cellular neural network (CNN), invented by Prof. Chua, et al., from the University of California at Berkeley, is a network consisting of a two dimensional array of locally interconnected analog processors. For image processing applications, each pixel of the image to be processed is usually associated with one cell. The processing is therefore fully parallel. Furthermore, as opposed to other neural network topologies, the interconnections between the processing elements are only local. This makes CNN quite suitable for VLSI implementation.

CNN has a wide range of applications, such as noise removal, smoothing, hole filling, edge, corner, shadow and motion detections. The application a CNN will serve depends on the values of the templates. The processing speed is very high. The typical settle time can be as small as 100 ns.

One limitation of early CNN invention is that it was "hardwired." That is, once you built it, it would perform one function and one function only. So if you wanted to perform a series of transformations, you would need a series of analog chips, one stacked on top of another. Chua et al. then invented a programmable CNN (or Universal CNN) which significantly improves analog computation. The chip could not only be programmed, but the results of each operation could be stored in the chip. Now a single "chip" could be programmed to perform one operation. Then the result of that operation could be stored in each element and the system reprogrammed to perform a second operation. The process could then be repeated many times. One CNN Universal chip could replace a series of hard-wired chips. The system is fast and efficient because the computation is analog. The system is extremely flexible because it can be programmed. It is easy to program because only local connections must be specified.
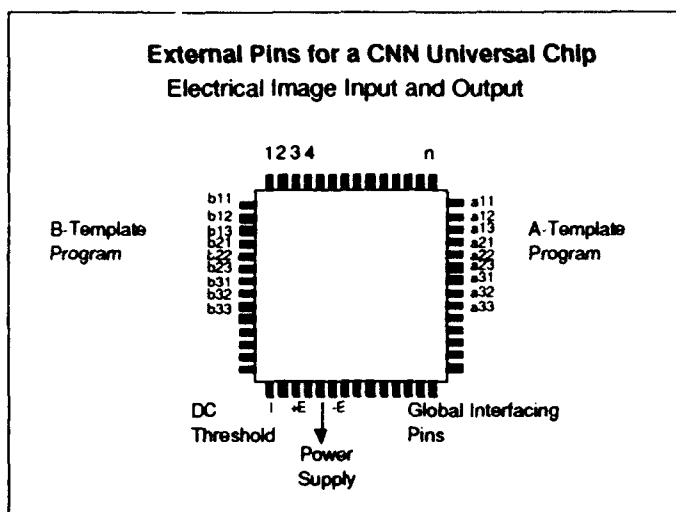
Figure 2-2
A diagram of the CNN chip. The pins at the top carry the image to be transformed.
The pins at the right and left, labeled "template," carry the transformation programs.

The CNN Universal Chip (Figure 2-2) is an extraordinarily fast, compact and powerful completely self-contained analog computer which can be built using currently available VLSI technology. It is capable of highly complex, real-time processing. The chip consists of an array of locally interconnected analog elements (cells), each with local analog and logical memory, but under global program control. It is programmable in a high-level language and includes its own compiler and operating system. Only 19 numbers are required to specify each processing operation, regardless of the size of the array. The CNN universal chip is particularly effective in processing two-dimensional patterns. Its logical capabilities make it possible for the chip to extract features of a "scene," and interpret (recognize) performance according to specified criteria. The CNN Universal Machine has the following characteristics:

*   Features of the CNN Universal Machine include: programming simplicity, a complex program (step-by-step sequence of instructions for implementing a flow chart or algorithm) can be written in a C-like language which is loaded, stored and executed within the chip through a CNN compiler and CNN operating system (which convert the program into internal macros and machine instructions to be implemented automatically like a Von Neumann machine). The local analog memory and logical memory within each cell allow this stored program to be executed (with additional communication and control circuitry

and clocks) without any off-chip communication during the execution. Only 19 real numbers (at external pins) are needed to program each operation. The chip is immediately applicable to digital systems with the integration of sensor arrays (on-chip sensors), an existing high level language for programming, and the integration of the output to digital microprocessors.

* Each CNN cell is a primitive analog computer: the CNN performs highly decentralized distributed computing with spatial and temporal computing units.

* Adaptability: the CNN can interpret environmental conditions and reprogram itself to optimize its operating characteristics for a particular environment. Both global and pixel by pixel adaptation are possible.

* Completely Internal Processing: no signal ever leaves the chip during the processing of a stored program. This eliminates extraneous noise, signal delay and input-output interface bottlenecks, as well as the need for external instructions during computation.

A CNN chip with a fixed template can perform about 10^{12} connections per second (XPS) using 2 μm technology on an area of 2 cm^2. The cost is only one cent per million operations per second. Each computational cell consumes just milliwatts. A CNN Universal chip with the same computing power can be built using 1.2 μm technology on an area of about 3 cm^2. Figure 2-3 compares a variety of different image processing devices on the market today. Their cost per million operations per second varies form $90 for the most expensive to $0.01 for the CNN, a difference of 4 orders of magnitude. In terms of speed, the range is from 600 operations per second to over 1 million operations per second for the CNN. Clearly in terms of speed and cost, the CNN is many orders of magnitude ahead of any of its competitors.
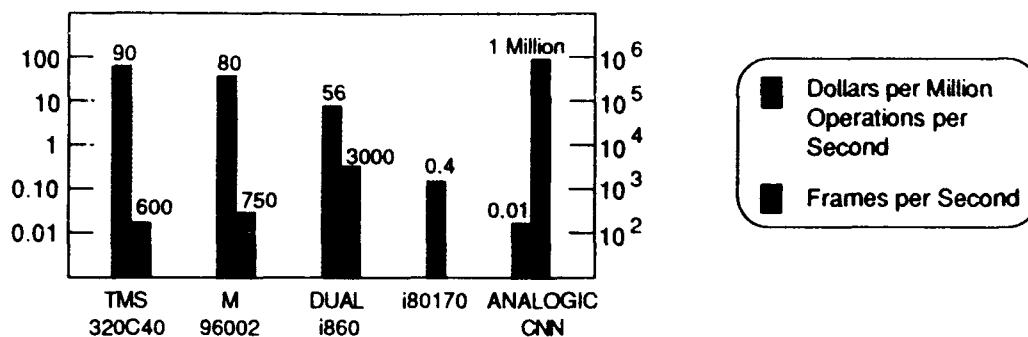
Figure 2-3
Graph showing the relationship between the costs of operation and frame processes per second
for five different image processing chips available today. The CNN is clearly better by many orders
of magnitude.

2.2   3-D Digital Image Processor Architectural Design

Figure 2-4 is a schematic diagram of parallel 3-D opto-electronic computer demonstration hardware which POC designed and constructed in one of its past government programs. This design was used in the present project to study the overall design trade-offs of the proposed 3-D hybrid cellular array machine. The hardware consists of three layers, each layer having direct communication paths with a host computer (HC), which in this case is a PC. Each layer (or board) contains four processor nodes, each node having its own signal processing circuit, and an optical interconnect interface circuit. The connections within a single layer are electrical, whereas the connections between planes are optical. Optical waveguide interconnects provide the means of interconnecting nodes on different layers. The data processing scheme is based on SIMD parallel pipelining. The PC is used to provide data, instructions, and synchronization to the various processor nodes.
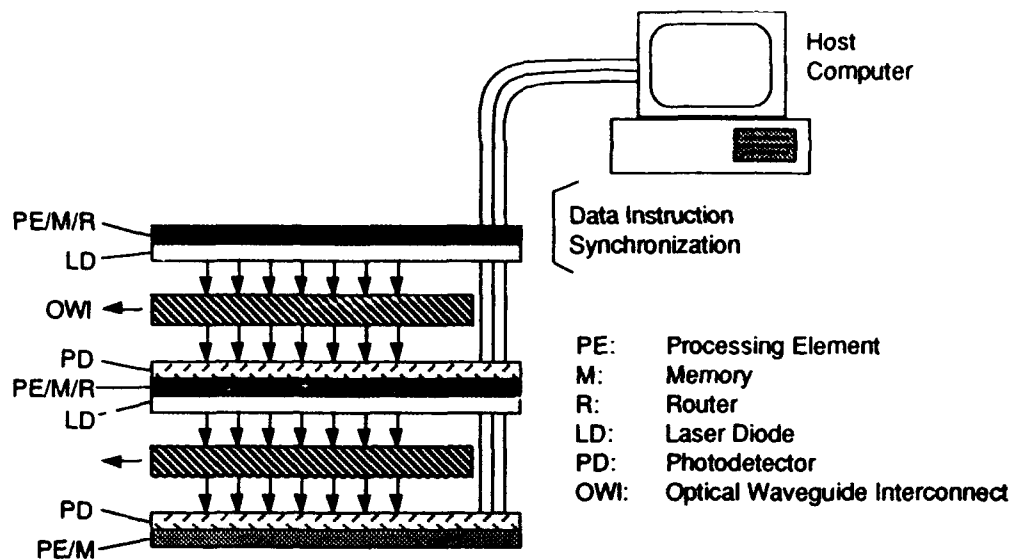
Host
Computer

PE/M/R
LD

OWI

PD
PE/M/R
LD

PD

PE/M

Data Instruction
Synchronization

PE: Processing Element
M: Memory
R: Router
LD: Laser Diode
PD: Photodetector
OWI: Optical Waveguide Interconnect

Figure 2-4
Schematic diagram of the setup of a 3-D opto-electronic computer.

Figure 2-5 shows a block diagram of the system design. Each board consists of four nodes, one interface to the PC, and four unidirectional optical interconnect channels to the next board with one channel per node. The nodes in the first layer only contain the laser diode driver circuit, and the nodes in the third layer only contain the photodetector receiver circuit. The nodes in the second layer have both circuits. Figure 2-6 illustrates the design of each layer. Two communication schemes are used. A parallel wrap-around mesh communication scheme is employed in the node array. A system bus is used to transfer data to/from the PC. This system bus solves the data transfer bottleneck between the PC and the data processing node. Because of this system bus design, the node will not be interrupted and can continue to process data and simultaneously transfer data to/from the HC. A dual port RAM in each node makes this uninterrupted data transfer possible. While the processing element/photonic interface units are performing their own operations, the data from the PC can be "silently" loaded into the RAM in each node without interrupting the processing element unit of the node. When the data transfer is completed, the direct memory access (DMA) in each node is activated and transfers data from the RAM to a local data memory in the processing element unit for the next cycle operation. Since these DMA transfers are effective for all nodes in the layer, the time spent in operation is reduced to a minimum and efficient pipelining is achieved.
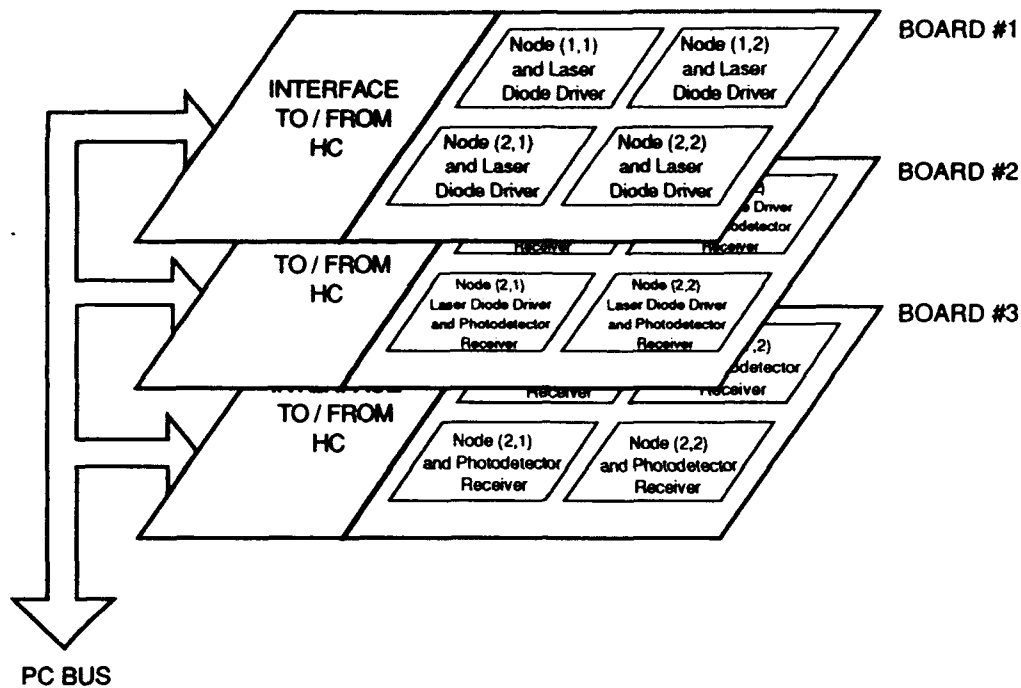
Figure 2-5
Design of each DSP layer.

Figure 2-7 describes the design of a single node. In this design, a Texas Instrument DSP chip (TMS320C31) is used as the processing element. The parallel fully interconnected scheme is supported by four bi-directional latches which provide communication to four nodes. A dual-port RAM (128K x 32) is employed with the system for uninterrupted DMA data transfer. The optical I/O interface is also shown in the figure. One output shift register, one input shift register, one laser diode driver, one photodetector amplifier/TTL converter, one laser diode, and one photodetector are incorporated. Each node consists of an emulator interface for final system hardware testing and debugging.
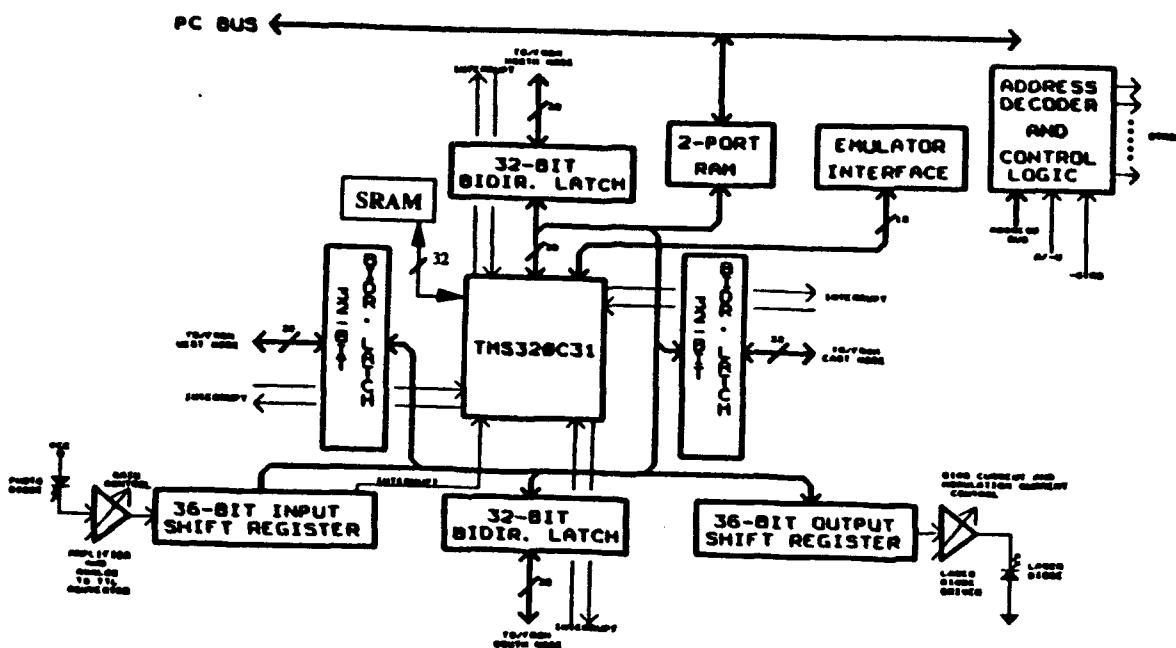
12

Figure 2-7
Design of a single node.

The communication scheme between the boards can be summarized as follows: board #1 can only transmit data to board #2 through optical channels (total of four optical diode transmitters). Board #2 can only receive data from board #1 and can transmit data to board #3 through optical channels (total of four optical receivers and four optical laser diode transmitters) and board #3 can only receive data from board #2 through optical channels (total of four optical receivers). Each optical transmitter converts 32-bit parallel data to a serial format, adds start and stop bits to total 36 bits, and transmits data serially. There is no interrupt on "transmit shift register empty" but only a status bit. Each optical receiver converts 36-bit serial data to a 32-bit parallel word (start and stop bits are thrown away) and generates an interrupt whenever a complete word is received. There is no error protection or error detection. The maximum serial transfer rate is 40 Mbs/sec.

Similarly, communication between nodes within a single board can be summarized as follows: there are four independent 32-bit bi-directional latches (one per pair of adjacent nodes). The transfer of data takes place one word at a time. Writing data to the latch automatically generates an interrupt to the destination node. Similarly, reading data from the latch automatically generates an acknowledge to the source node. There is a common system clock for all three boards and the

clock signal is distributed electrically (there is no clock signal encoded in the data during optical transmission).

The processing power of the entire system is based on the performance of the TI DSP chip (TMS320C31). The chip has the following performance: 60 ns single-cycle instruction execution time, 40 MFLOPS and single-cycle multiplication/accumulation operation. Features include separate program/data/DMA buses, two on-chip 1K x 32-bit RAMs, one 64 x 32 bit program cache and a 16 M word external address register. The DSP chip can execute 32-bit floating-point multiplication and ALU operations in a single cycle (60 ns).

Photographs of the complete hardware system are shown in Figure 2-8. The results of the architectural design study for the proposed 3-D hybrid cellular array machine based on the above constructed hardware are summarized in Table 2-1.

It was found that the opto-electronic interface chip in the 3-D opto-electronic computer project was the bottleneck of system performance. In Phase II, we will design a single chip for the opto-electronic interface. It will have a much higher data rate and an error detection/correction capability. With this new design, the processor nodes between layers can maintain the same data communication bandwidths as the processor nodes within each layer. In other words, for a 32-bit processor node, >40 Mbits/sec per bit throughput can be achieved (32 bit x 40 Mbits/sec bit = 1.28 Gbits/sec). Section 3.0 details the design of this opto-electronic interface chip.

To summarize, the feasibility of the 3-D cellular array machine has been proven since a similar system has already been fabricated and demonstrated by POC. The design issues listed in Table 2-1 were identified as key elements in developing the proposed 3-D hybrid cellular array machine in this program for real-time image processing.
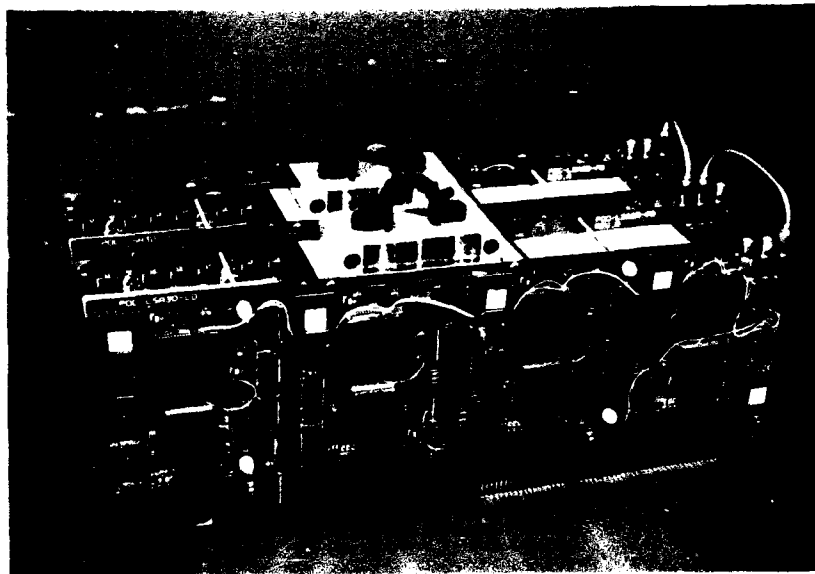
(a)

(b)

Figure 2-8
3-D opto-electronic computer hardware: (a) the 3-D opto-electronic computer and (b) the entire
system, including a host computer (PC), a display module, the 3-D opto-electronic computer in an
industrial computer box and debugging hardware.

Table 2-1 Improved Architectural Design

	3-D Opto-electronic Computer (Constructed Hardware)	3-D Hybrid Cellular Array Machine (Planned Hardware Design)
Processing Techniques	Digital Only	Digital and Analog
Analog Processor	N/A	Cellular Neural Network Chip
Digital Processor	TI C31	TI C40
Optical Interconnect	40 Mbits/sec per node	1.0~1.5 Gbits/sec per node
Opto-electronic Interface Chip	Multiple Discreet Electronic Components No Error Detection/Correction	Single Chip Design Error Detection/Correction

2.3 System Simulation and Verification

In order to understand the performance of the 3-D hybrid cellular array machine designed in this project, a computer simulation program was written to simulate the cellular array machine's architecture for processing different object images using both analog and digital processing techniques. Computer simulation was performed on images obtained from a frame grabber. The goal was to demonstrate how digital VLSI image processing techniques can overcome the precision problem of analog processing and still offer high processing speed by only handling the critical regions of the incoming object images.

The computer simulation performed in Phase I is illustrated in Figure 2-7. A highly noisy, low-contrast input is first presented to an analog image processor for edge enhancement. The processed image data, together with the original raw image data, is fed into a digital image processor for further processing. In this step, the digital image processor does not have to process the entire frame of image data. Based on the preprocessed image data and the coordinates of the critical regions provided by the analog image processor, the digital image processor needs only to process, in a highly precise manner, these critical regions.
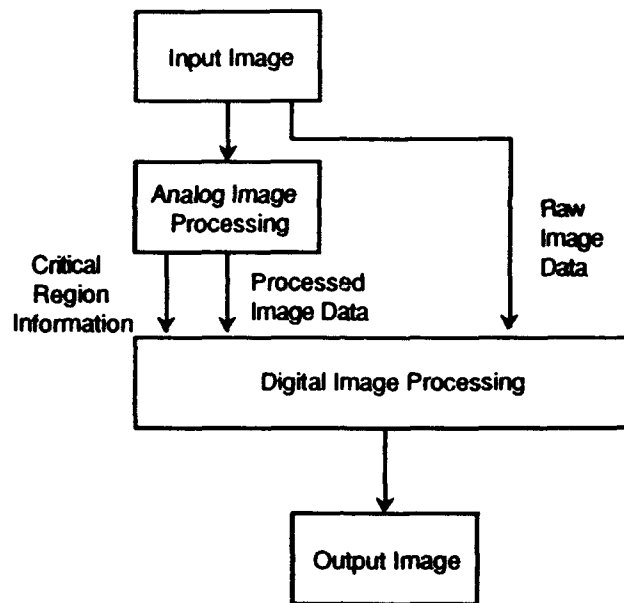
```
            ┌─────────────┐
            │ Input Image │
            └─────────────┘
```

Figure 2-7
Flow chart of the Phase I computer simulation.

Analog Processing Simulation

In this simulation, we tried to prove that combining both analog and digital image processing techniques will result in a better performance, not only in terms of processing speed but also in computation precision. Three images were used in this simulation. Images were first grabbed by a frame grabber. The gray level of the frame grabber was from 0 to 256, i.e., 8 bits. Different degrees of additive uniform or Gaussian noise were added to the images. A common characteristic of the three images was low luminance. As mentioned above, analog circuits generally have a smaller dynamic range. In our program, we converted the gray level of the input image to a smaller range (4 ~ 6 bits). Pixels whose gray level was higher or lower than the chosen range were considered saturated and were given maximum and minimum values after conversion. Then, the image was filtered by a low pass filter. The low-pass template used was a 3×3 matrix defined as

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

17

The edge of the image was then detected. The edge-detection templates used were

$$\begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}.$$

For both low-pass filtering and edge detection, we used only locally connected cell information. Since the images we used had low luminance, the output of the edge detection, as expected, was poor. Some edges were missed because of the saturation effect. The simulation then detected the regions of the processed image which were incomplete and determined the regions (i.e., critical regions) which required further processing.

Digital Processing Simulation

After the analog pre-processing, both processed image data and raw image data were sent to the digital processor. The digital processor reprocesses the critical regions from the raw image data. In our simulation, both digital median filtering and morphological processing techniques were used in these pre-processing operations.

Figures 2-8(a), 2-9(a) and 2-10(a) show the input raw images we used in the simulation. Figures 2-8(b), 2-9(b) and 2-10(b) show the processed image data after processing through the analog image processor at different precision levels. The precision level of the analog processor is between four- and six-bit accuracy. Figures 2-8(c), 2-9(c) and 2-10(c) illustrate the final image data obtained from the digital image processor. Only several critical regions were selected for processing. An eight-bit accuracy median filter and morphological processing operation was used in the digital image processor. It is clear that the image edges are much more complete when compared with the output from the analog image processor alone.
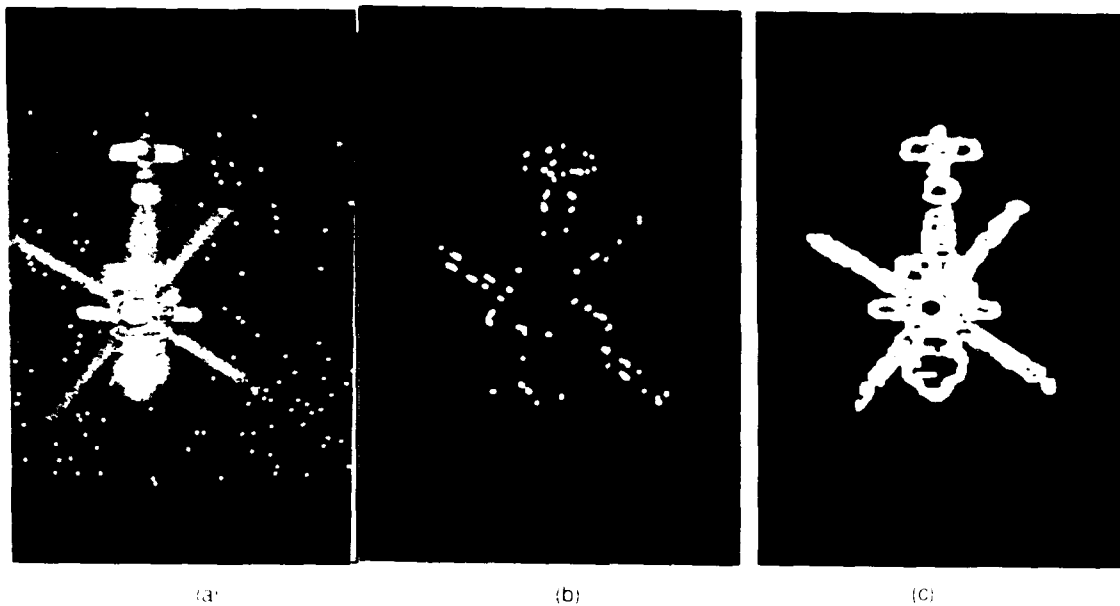
Figure 2-8
Computer simulation result #1 (a) input raw image data, (b) processed image data from an analog image processor (with a four bit accuracy) and (c) final output image data after both analog and digital image processing
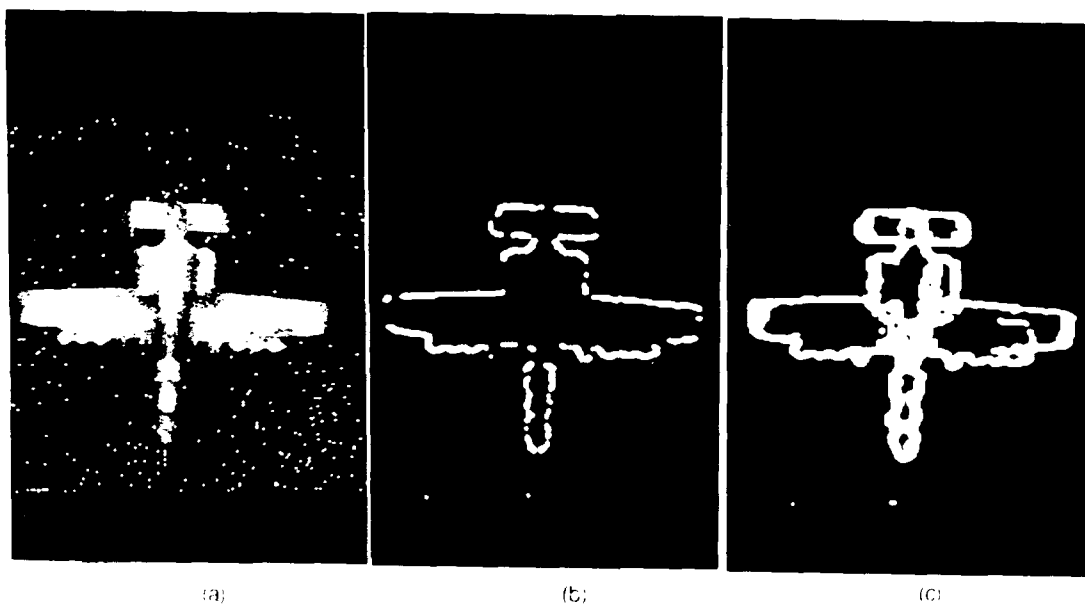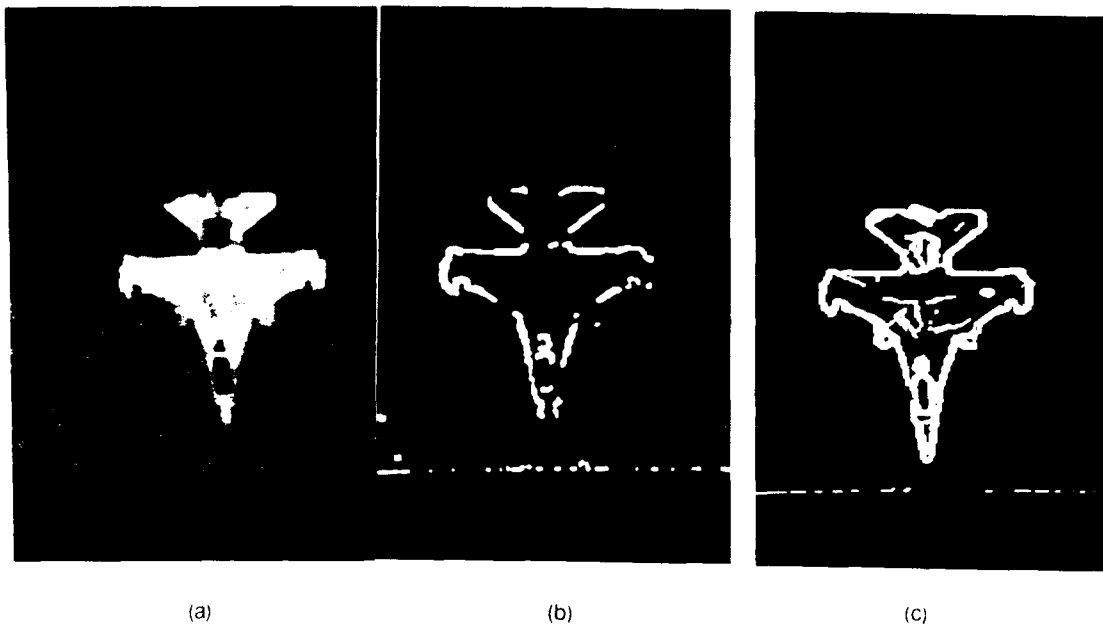

Figure 2-9
Computer simulation result #2 (a) input raw image data, (b) processed image data from an analog image processor (with a five bit accuracy) and (c) final output image data after both analog and digital image processing
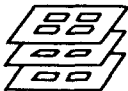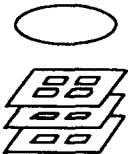
(a) (b) (c)

Figure 2-10
Computer simulation result #3: (a) input raw image data, (b) processed image data from an analog
image processor (with a six-bit accuracy) and (c) final output image data after both analog and
digital image processing.

Table 2-2 lists various image processing architectures. Row 1 shows the analog image processing technologies. They are high-speed and compact. But they are low-precision and hard-wired. Rows 2, 3 and 4 show the digital processor architectures. They are high-precision and programmable. But the speed is limited and the system size is large. Row 5 shows an architecture which uses analog preprocessing and a 3-D cellular array machine. This last design retains most of the advantages of both the analog and digital designs, such as high-speed, high-precision and programmability.

Table 2-2   Comparison of Various Architectures for Image Processing

		Architecture	Performance	Characteristics
1		Analog Image Processing (Cellular Neural Network or Silicon Retina)	10^{12} XPS/cm^2	+ High Speed + Compact Size + Low Power Consumption - Low Precision - Hardwired or Limited Algorithm Efficiency - Limited Programming Flexibility
2		Single Computer From PC, DSP to Supercomputer	100 MOPS (DSP assumed)	+ High Precision
3		Digital Cellular Array Machine Assuming an M x M Array	M x M x 100 MOPS	+ Highly Efficient Programmability + High Level Processing
4		3-D Digital Cellular Array Machine Assuming N Layers of M x M Arrays	N x M x M x 100 MOPS	- Limited Speed - High Power Consumption
5		3-D Digital Cellular Array Machine with Analog Preprocessing	N x M x M x M x 100 MOPS + 10^{12} XPS	Retains most of the advantages of both analog and digital processing, such as high-speed, high-precision and programmability. The size and power consumption of the digital unit can be reduced because most of the low-level processing is performed by the analog unit.

3.0   OPTICAL TRANSCEIVER CHIP DESIGN

As described in Section 2.2, a single chip implementation of the photonic interface function is required in the proposed optically-assisted 3-D hybrid cellular array machine. Figure 3-1 shows a block diagram of an optical transceiver design. In this optical transceiver chip, the 32-bit data bus is used to communicate with the host machine, while one 8-bit local bus is used for the optical receiving channels and another 8-bit data bus for the optical transmitting channels. The chip consists of an I/O interface, a multiplexer (MUX), a demultiplexer (DEMUX), an optical transmitter (TX) and an optical receiver (RX). The CPU block represents the digital host machine. Each 32-bit data word from the host machine is segmented into 4 bytes and every byte is sent to the

transmitter section to be transmitted through the laser diodes. At the receiving end, incoming optical signals are detected by the p-i-n diode array and demodulated into digital signals by receiver circuits. Four bytes of data are combined through the multiplexing circuitry to form a 32-bit word and are sent to the digital host machine. To achieve the desired 1 Gbit/s operation, the individual transmitter and receiver circuits should operate at 125 Mbit/s and the clock rate of the digital host machine should be higher than 31.25 MHz. A detailed schematic of the I/O interface is given in Figure 3-2.
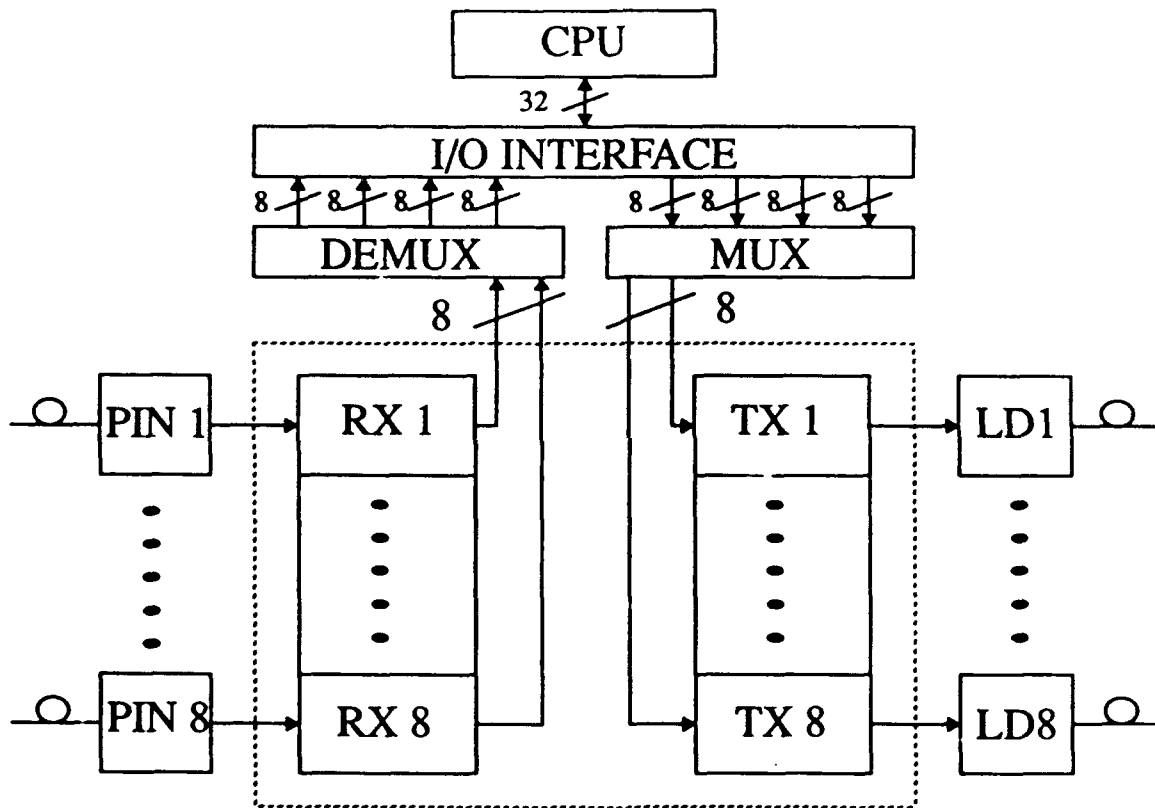

Figure 3-1
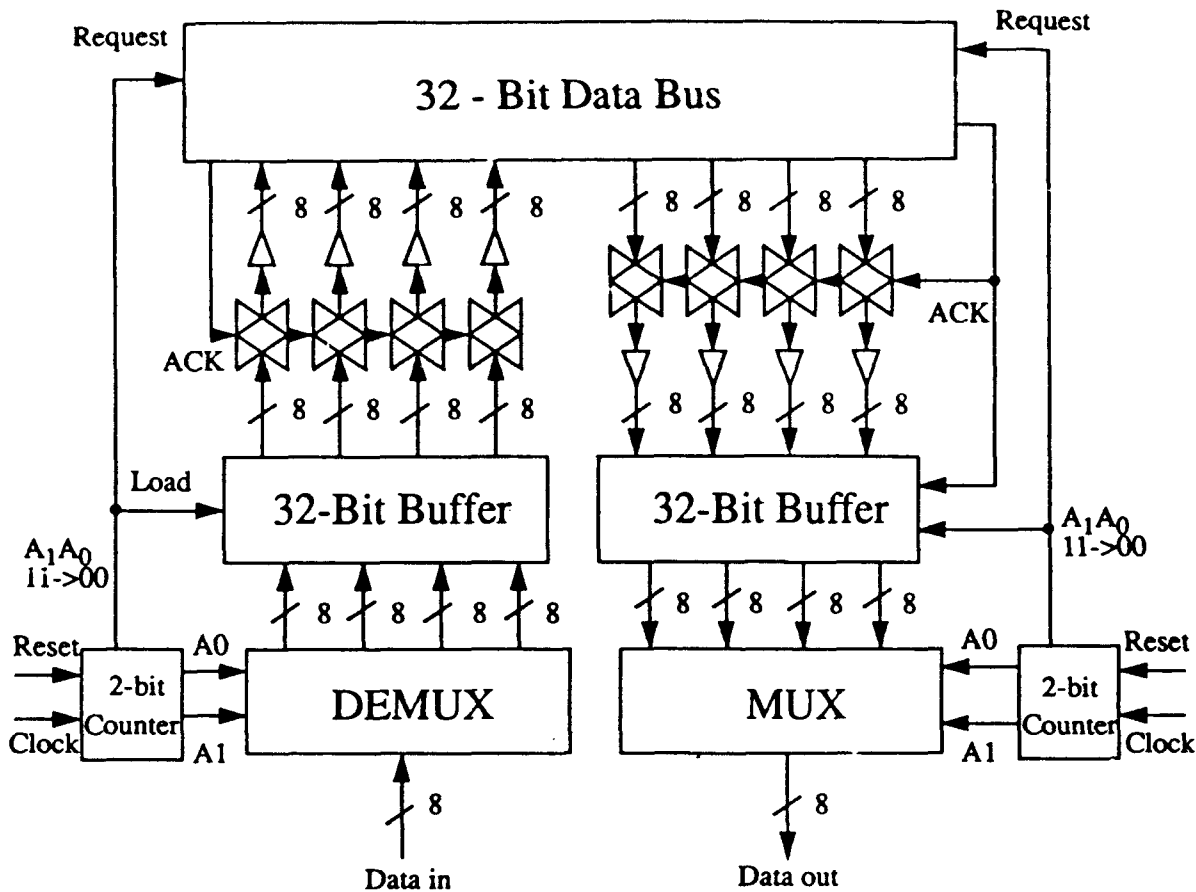Block diagram of the optical interconnection system.

Figure 3-2
Block diagram of the I/O interface.

Figure 3-3 shows a typical block diagram of the receiver circuit. It consists of a preamplifier, automatic-gain-control amplifier, decision circuit and clock recovery blocks. The clock recovery circuit can be omitted for short-distance communication. The incoming optical signals are detected and converted into electrical current signals by the photodetectors, such as the p-i-n diode. Characteristics of various commercially available p-i-n diodes are listed in Table 3-1. For example, the responsivity of the model PIN-HS008 device at the 830-nm wavelength is 0.4 A/W, and its response time, junction capacitance and dark current at -10 V are 1 ns, 1.5 pF, and 1 nA, respectively. The electrical current signals are amplified by the amplifier chain, which includes a low-noise preamplifier and an automatic-gain-control main amplifier. The decision circuit samples the signal and provides the binary outputs by the thresholding function. The clock recovery circuit extracts the clock which is used for sampling and further operation.
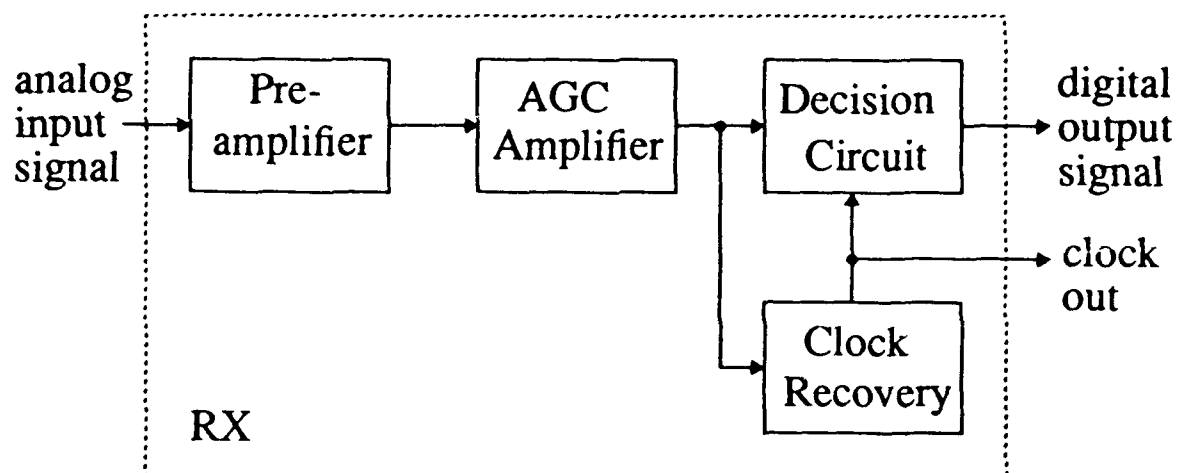
Figure 3-3
Block diagram the receiver module.

Table 3-1   Characteristics of Various PIN Diodes

Model #	Responsivity		Response time		Capacitance	Dark Current	
	@830 nm		@-10 V		@-10 V	@-10 V	
	R (A/W)		t_r (ns)		C_j (pF)	I_d (nA)	
	Min.	Typ.	Typ.	Max.	Typ.	Typ.	Max.
PIN-HS008	0.3	0.4	0.35	1.0	1.5	0.02	1.0
PIN-HS040	0.3	0.4	0.8	1.0	5.5	0.25	1.0
PIN-HR008	0.48	0.52	1.0	3.0	1.5	0.03	1.0
PIN-HR040	0.48	0.52	1.0	3.0	5.5	0.30	2.0

24

A block diagram of the transmitter circuit is shown in Figure 3-4. It contains a driver circuit and a current source. The driver circuit converts digital binary signals into laser-diode driving currents. The laser diode converts the driving currents into optical power output which is transmitted through the optical fiber. Characteristics of various commercially available laser diodes are listed in Table 3-2. For example, the optical power output of the model LT022PS device is 5 mW, its operating temperature is between -10°C and 70°C, and its threshold current is 45 mA. Its wavelength, operating current, and operating voltage at 3 mW optical power output are 780 nm, 55 mA, and 1.75 V, respectively. The current source provides two driving currents, I_{on} and I_{off}. In some approaches, the current source can be controlled by a proportional-to-absolute-temperature (PTAT) current reference which provides the positive temperature coefficient of current necessary to cancel the negative temperature coefficient of emitted optical power.
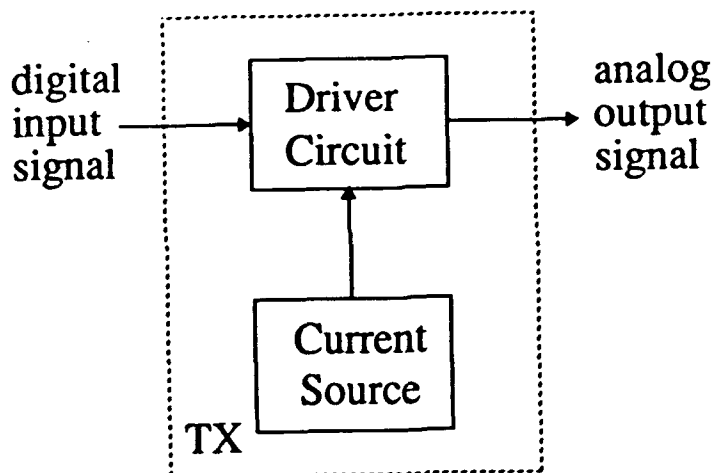

Figure 3-4
Block diagram of the transmitter module.

25

Table 3-2   Characteristics of Various Laser Diodes

Model #	Optical power output	Operating temperature	Wavelength			Threshold Current		Operating Current		Operating Voltage	
			@P_0 = 3 mW					@P_0 = 3 mW		@P_0 = 3 mW	
	P_0 (mW)	T_{opr} (°C)	λ_p (nm)			I_{th} (mA)		I_{op} (mA)		V_{op} (V)	
			Min.	Typ.	Max.	Typ.	Max.	Typ.	Max.	Typ.	Max.
MLA102A MLA402A MLA412A	5	-40 ~ +60	765	780	795	40	60	50	70	1.8	2.5
LT022HS	5	-30 ~ +85	770	780	795	45	70	55	85	1.75	2.0
LT022PS	5	-10 ~ +70	770	780	795	45	60	55	75	1.75	2.0
LT022MS	5	-10 ~ +60	770	780	795	45	70	55	85	1.75	2.0
LT022MC LT022MD LT022MF	5	-10 ~ +60	770	780	790	40	8	65	100	1.75	2.2

Since eight copies of the transmitter circuit and eight copies of the receiver circuit will be integrated into a single IC chip, special care should be paid to maintain a stable operation of the receiver circuit. A large signal swing in the transmitter circuit might perturb the receiver circuit by means of coupling through the substrate [15,16]. Coupling noise between transmitter circuit and receiver circuit should also be analyzed and carefully controlled. Table 3-3 is a summary of the results found in our design study of the Phase I work.

Table 3-3   Summary of Results from Preliminary Study

Optical Receiver			Transmitter			
Technology	1.2 μm CMOS	0.8 μm CMOS	Technology		1.2 μm CMOS	0.8 μm CMOS
Clock	83 MHz	125 MHz	Clock		83 MHz	125 MHz
Rise time	4.2 ns	2.8 ns	Rise time		2.9 ns	1.9 ns
Fall time	4.3 ns	2.9 ns	Fall time		2.4 ns	1.6 ns
No. of transistors	40	40	No. of transistors	Scheme 1	12	12
				Scheme 2	35	35
Dark current	1 nA	1 nA	Stand-by current		4.9 mA	4.9 mA
Modulation current	3 μA ~ 10 μA	3 μA ~ 10 μA	Drive current		62 mA	62 mA
Output signal level	CMOS: 0 V, 5 V	CMOS: 0 V, 5 V	Optical power output @ 25 °C		2.5 mW	2.5 mW

3.1   Optical Transmitter

A schematic diagram of the transmitter circuit is shown in Figure 3-5 [17]. 0.8 μm CMOS technology was used in the preliminary design and simulation. Transistor M9 is driven by an inverter chain to guarantee that the structure can be driven by normal logical circuits. This transistor is switched between the on state and the off state by the digital input signals. In the design and analysis, the width and length of transistor M9 were determined to be 320 μm and 3.2 μm, respectively. The low-level current through the laser diode is generated by transistor M10.

The transmitter circuit was simulated by the SPICE circuit simulator with the widely used BSIM (MOS level-4) model. The laser diode was simulated by an equivalent circuit model which consisted of a resistor and a voltage source in parallel with a capacitance. Figure 3-6(a) shows a plot of the output-driving current versus input data voltage. If the input voltage is smaller than 2.5 V, the output driving current remains at 4.96 mA, the stand-by current to reduce the transition time between the low-level input signal and the high-level input signal. The ON current is

62.5 mA for the high-level input signal. The voltage waveform at the output node is also shown in Figure 3-6(b). Figure 3-6(c) shows the transient simulation results of the transmitter circuit for the 200 MHz nonreturn-to-zero (NRZ) data. The input binary data is set to 1010. The rise time is the time for the transmitter to generate the high-level driving current when the binary input is 1. Similarly, the fall time is the time for the transmitter to generate the low-level stand-by current when the binary input is 0. The simulated rise time and fall time were 1.9 ns and 1.6 ns, respectively. The transient simulation results of the transmitter circuit for 125 MHz return to zero (RZ) data are shown in Figure 3-6(d). The input binary data is set to 1101. Note that the 0.8 μm CMOS technology is adequate for the integration of multiple transmitter circuits in the parallel data link application.

Figure 3-5
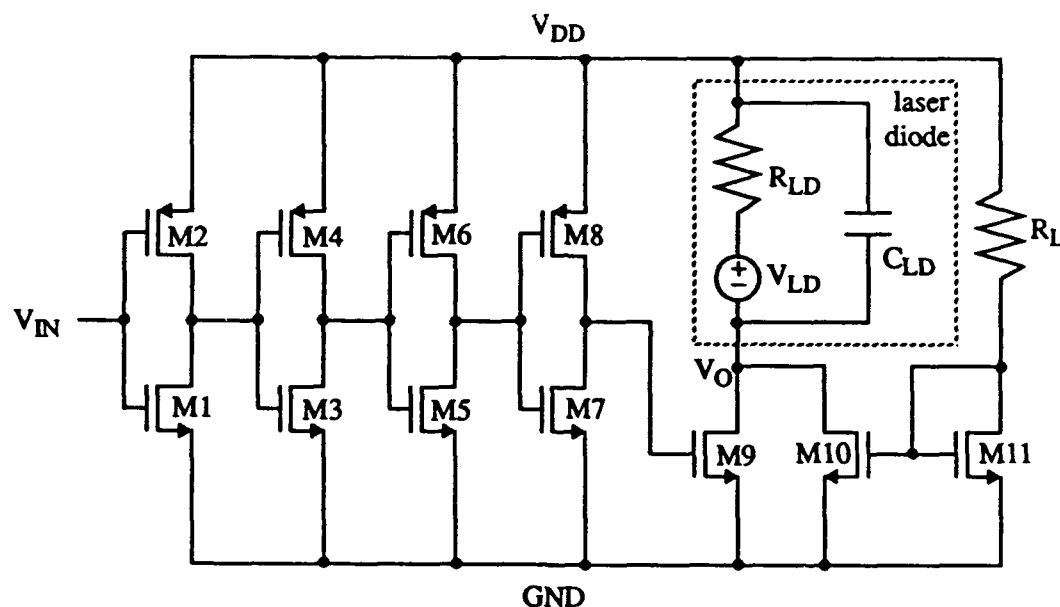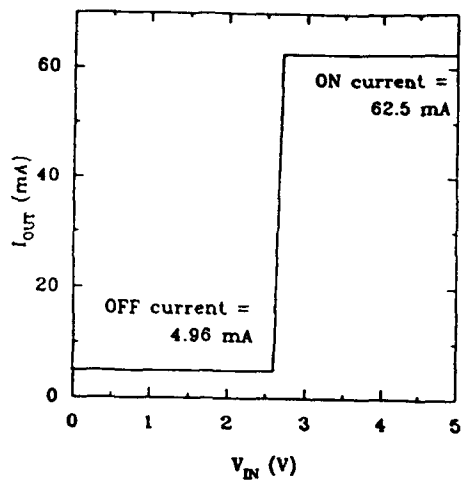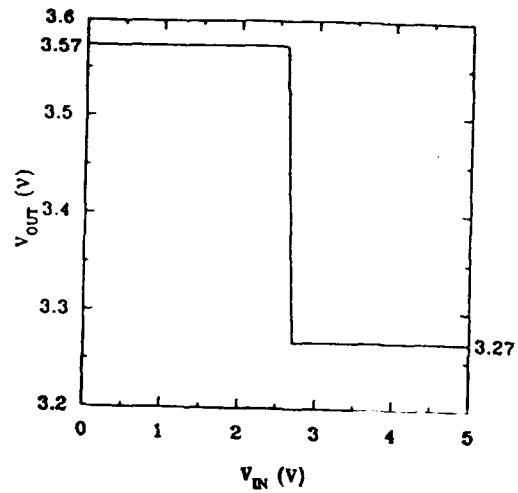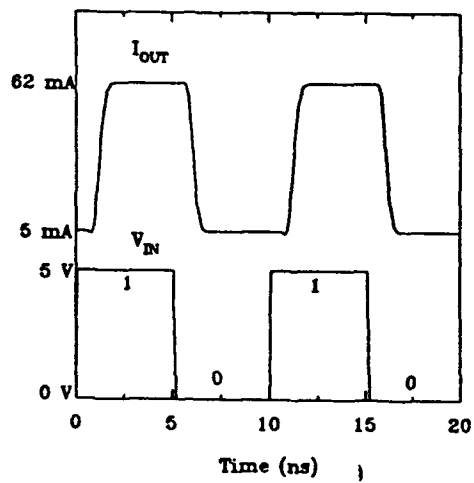**Schematic of transmitter circuit.**
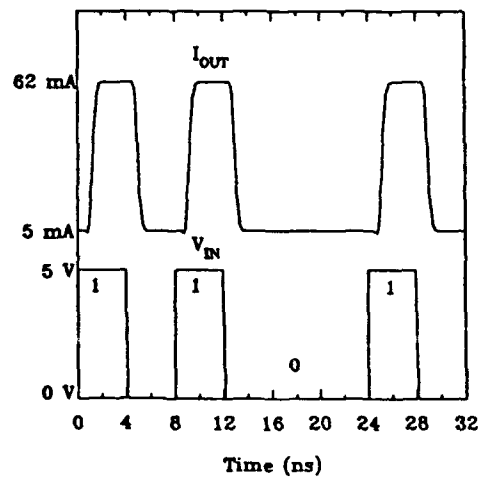
Figure 3-6
(a) Output driving current versus input voltage. (b) Output voltage versus input voltage.
(c) Transient simulation for 200 MHz NRZ data. (d) Transient simulation for 125 MHz RZ data.

The control of bias voltage of the laser diode to compensate for changes in threshold is another design issue that needs to be carefully considered. Without compensation, a constant bias can result in unacceptable changes in output power, extinction ratio, and turn-on delay. Figure 3-7 shows the circuit schematic of a modified transmitter circuit with a negative feedback regulator. The mean-power reference level and peak driving current are initially adjusted at 25°C to set the power level while optimizing extinction ratio and turn-on delay. The photodiode that monitors the light output is usually placed in the laser package to intercept the back-mirror output. For example, the ML4xx2A-series AlGaAs laser diodes are hermetically sealed devices having a Si photodiode for monitoring the light output. Output current of the photodiode can be used for automatic control of the operating currents or case temperatures of the lasers.
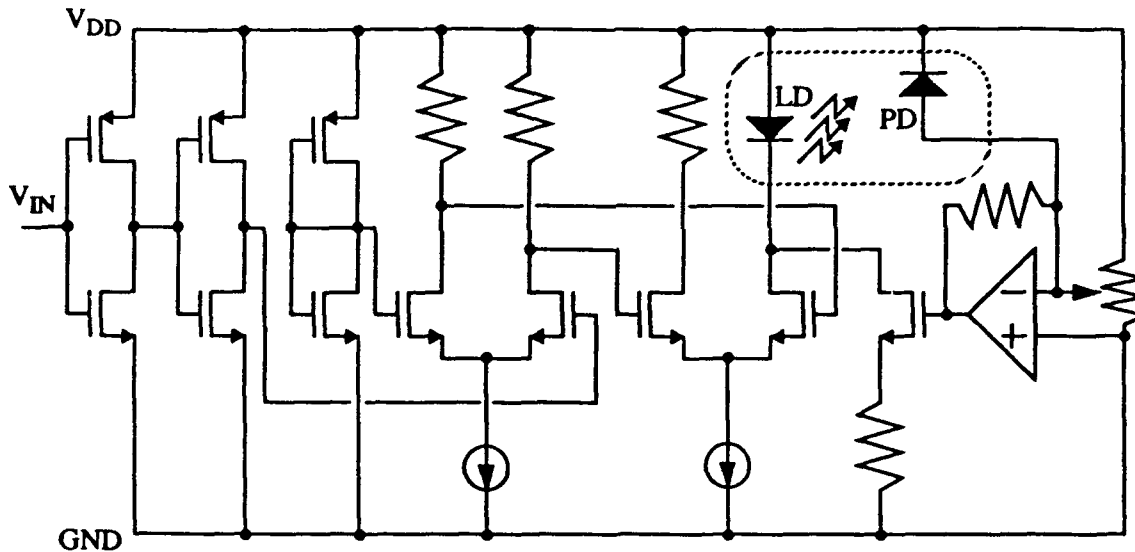

Figure 3-7
Schematic of the transmitter circuit with a negative feedback regulator.

3.2 Optical Receiver

The schematic diagram of one receiver circuit is shown in Figure 3-8 [17]. It is a three-stage amplifier, followed by a string of inverters. The basic operation of the amplifier is current-voltage-current conversion. However, the last stage only performs the current-to-voltage function. After the three-stage amplification, the signal strength was still not great. Thus, an inverter chain was

added to amplify the signal even further and to perform pulse shaping and buffering. Changing the current through transistor M9 can shift the DC level at the input of the first inverter to be equal to the threshold of the inverters. The circuit was designed with 0.8 μm CMOS technology and contains only n-channel MOS transistors in the signal path. In addition, all MOS transistors are designed and optimized to ensure that the transistors are always in saturation. This technique is the key to achieving the high switching speed. Current sources can be constructed of p-channel MOS transistors and can completely control the biasing of the circuit.

The receiver circuit was simulated by the SPICE circuit simulator with the BSIM model. The photodiode was represented by a parallel combination of the signal current source, the dark current source and the capacitance under the operating bias voltage. Figure 3-9(a) shows the plot of the output voltage versus the input current. If the input current is greater than 5 μA, the output voltage is 5 V, which is the high-level state of CMOS circuits; otherwise, the output voltage is 0 V, which is the low-level state of the CMOS circuits. Figure 3-9(b) shows the transient simulation results of the receiver circuit for the 200-MHz NRZ data. The simulated rise time and fall time were 2.8 ns, 2.9 ns, respectively. The transient simulation of the receiver circuit for the 125-MHz return-to-zero (RZ) data is shown in Figure 3-9(c).

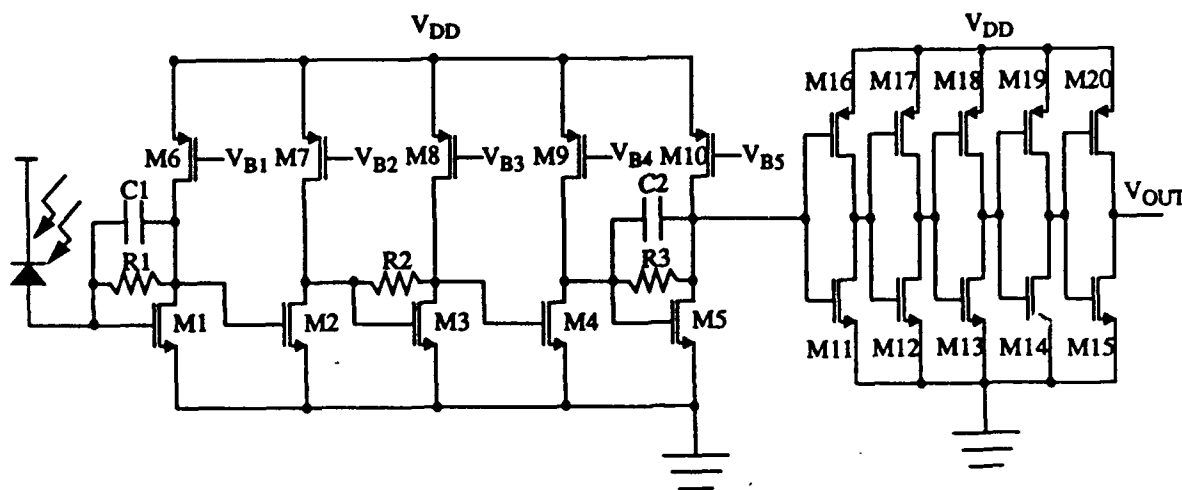Figure 3-8
Schematic of the receiver circuit.
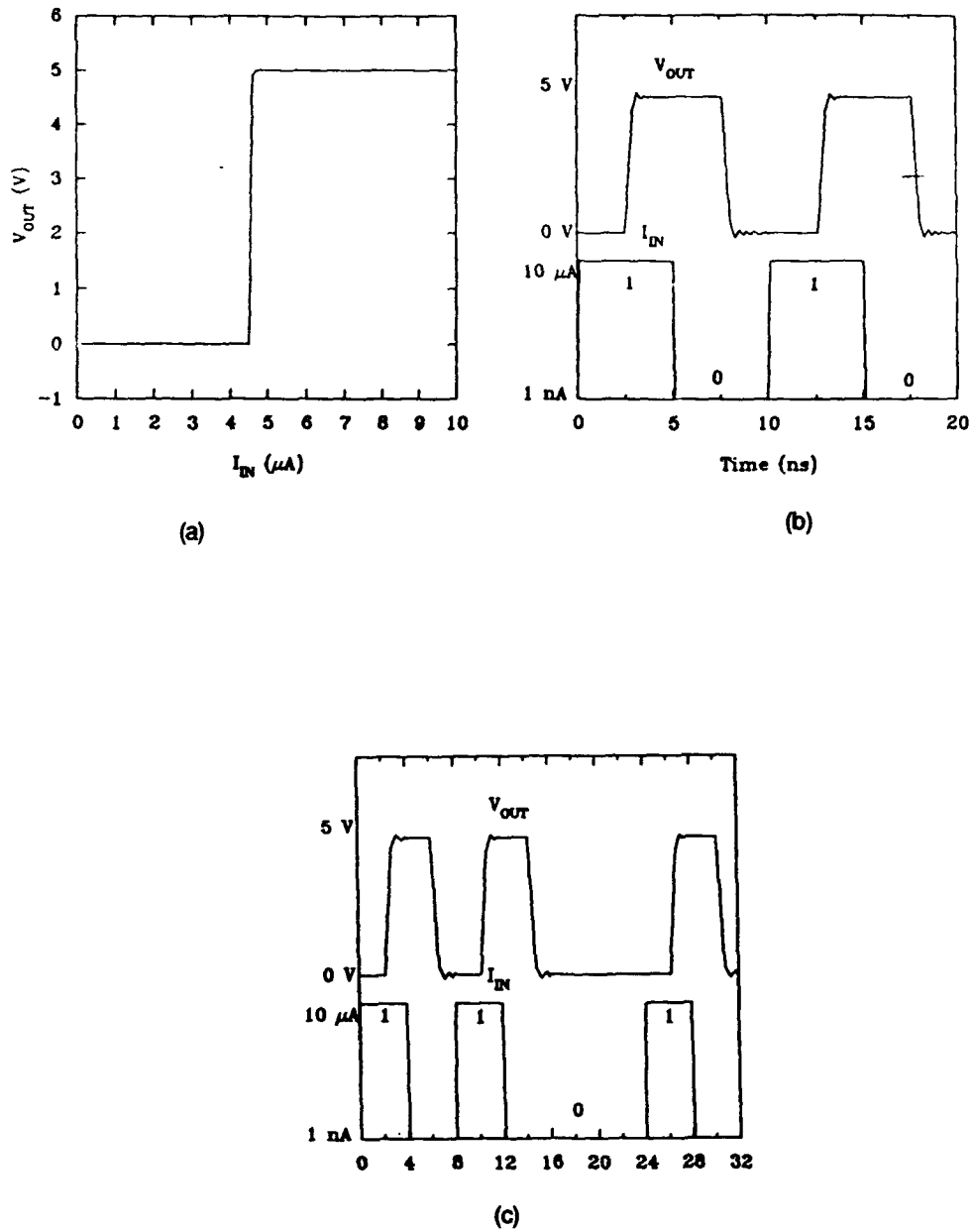
(a)

(b)

(c)

Figure 3-9
Simulation results of receiver circuit. (a) Output voltage versus input circuit. (b) Transient
simulation for 200 MHz NRZ data. (c) Transient simulation for 125 MHz RZ data.

Receiver noise is an important factor in determining receiver sensitivity. Receiver noise can be represented by an equivalent noise current source at the input node. For the field-effect transistor (FET) front-end preamplifiers, the square input equivalent noise current operating at bit-rate B can be expressed as [18]

$$< i_n^2 > = 2q(I_D + I_I)I_2 B + \frac{4kT}{R_{FB}}I_2 B + \frac{4kT}{g_m}\Gamma(2\pi C_T)^2 f_c I_f B^2 + \frac{4kT}{g_m}\Gamma(2\pi C_T)^2 I_3 B^3. \quad (3\text{-}1)$$

Here, q is the electron charge, k is Boltzmann's constant, T is the absolute temperature, I_D is the dark current of the photodiode, I_L is the leakage current of the transistor, g_m is the transconductance parameter of the transistor, Γ is the transistor noise figure, C_T is the total capacitance at the input node, f_c is the 1/f noise corner-frequency, I_2, I_3, and I_f are the effective receiver bandwidth integrals. Figure 3-10 shows the relationship between the square input equivalent noise current and the data bit-rate B. As the bit rate increases, the contribution from the leakage current decrease while the channel noise and 1/f FET noise become dominant. In addition, the total square noise current is proportional to C_T^2, which is again the critical design factor for low noise operations. The receiver sensitivity expression is [18]

$$\eta \overline{P} = Q\frac{h\Omega}{q}\sqrt{< i_n^2 >} \quad (3\text{-}2)$$

where $h\Omega$ is the photon energy ($= hc/\lambda$) and Q is a parameter relating to the desired error rate. In our design and analysis, the square input equivalent noise current operating at the bit-rate 125 Mbps was 6.6×10^{-16} A^2, which is equivalent to 25.7 nA-rms. The ideal receiver sensitivity, without device mismatch effects, as predicted by Eq. (3-2) is -63 dBm. The expected measured sensitivity with device mismatch effects will be in the range of -14 dBm to -20 dBm.
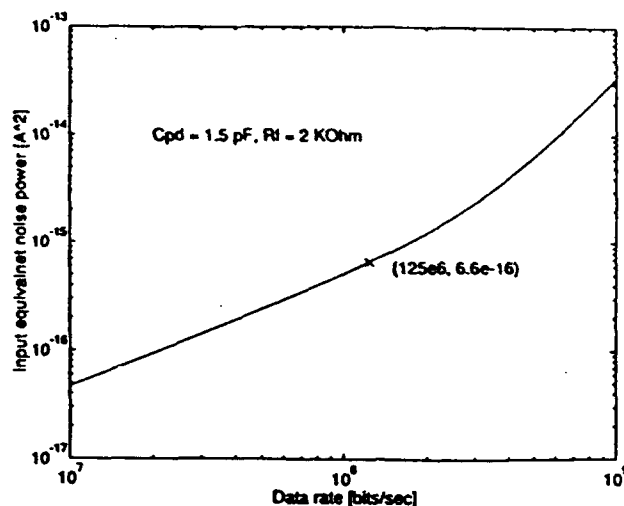
Figure 3-10
Relationship between the noise performance and the data rate (calculated result).

3.3　　Summary

The integrated transceiver chip for the optical fiber data link will include a 32-bit data bus to interface with the digital host machine, an 8-bit local bus for the receiving channels, and another 8-bit local bus for the transmitting channels. The 0.8 μm CMOS technology from the MOSIS Service of USC/Information Science Institute is adequate to support the design and fabrication of the integrated optical transceiver chip at low manufacturing cost. Several major engineering design challenges still need to be carefully addressed in optimizing the performance of the optical data link chip. Electrical crosstalk between transmitter circuits and parallel receiver circuits is critical and must be minimized. Extensive design, detailed circuit-level simulation, and careful layout of the essentially analog circuits of the transmitter and receiver sections will demand significant engineering effort and resources. Our Phase II work will continue this effort.

4.0　　CONCLUSIONS

A preliminary design of a real-time image processing system based on the proposed optically-assisted 3-D hybrid cellular array machine was produced. In addition to 3-D digital processor arrays, the system incorporates analog front-end units to accelerate the processing. These analog

34

units perform fast, low level processing, while digital units carry out high precision calculations on critical regions (such as corners, beginning and end points of lines, or broken edges). Further, the digital processors can also be used for high-level processing and executing algorithms that cannot be done by analog units. The system utilizes optical interconnects to provide highly efficient, high-speed communication paths to link the digital processor layers and to connect the analog unit with the digital processor layer.

Conclusions drawn from the Phase I work are that

1. Cellular Neural Networks and the CNN Universal Machine were investigated as the analog processing elements. These devices are very fast. A large number of templates cover a wide range of low level processing. Small scale prototype chips have proven feasibility of the architecture. Each cell is a simple circuit. Duplication of those cells can yield a VLSI chip.

2. Digital signal processors with nearest neighbor interconnects are the engines for further parallel processing. Layers of boards containing digital arrays can be stacked together for pipeline and/or concurrent processing. The need for the digital processor are evident: analog elements are generally imprecise and inflexible. The CNN Universal Machine incorporates a certain degree of programmability. Experiments have shown that the CNN is capable of achieving good results when the image with high SNR is present. For gray-level images, in particular, for low-contrast images in high clutter, the imprecision of the analog circuits sets the limitation of processing accuracy. Our computer simulation work has verified this limitation. Digital processing layers will be needed for high precision processing. Algorithms that cannot be easily executed by the CNN can also be handled digitally.

3. Parallel and high bandwidth optical interconnects are needed to maintain communication among both the analog and digital processing layers. For the opto-electronic (photonic) interface chip, 0.8 μm CMOS technology from the MOSIS service is adequate. A number of design issues, such as crosstalk through the power and ground lines, need to be carefully addressed.

5.0   REFERENCES

1.   Ngan, K. N., Kassim, A. A., and Singh, H. S., "Parallel Image Processing System Based on the TMS32010 Digital Signal Processor," Proc. IEEE, Vol. 134, Part E, No. 2, pp. 119-124, March 1987.

2.   Vemuri, Ed., "Artificial Neural Networks: Theoretical Concepts," New York: IEEE Computer Society Press, 1986.

3.   Mead, C., "Analog VLSI and Neural Systems," Chap. 15, pp. 257-278, Addision Wesley, 1989.

4.   Chua, L. O., and Yang, L., "Cellular Neural Networks: Theory," IEEE Trans. Circuits Syst., Vol. CAS-35, pp. 1257-1274, Oct. 1988.

5.   Baktir, I., and Tan, M., "Analog CMOS Implementation of Cellular Neural Networks," IEEE Trans. Circuits Syst., Vol. 40, No. 3, March 1993.

6.   Chua, L. O., and Yang, L., USA Patent No. 5140670, August 18, 1992.

7.   Chua, L. O., and Roska, T., "CNN Universal Machine and Supercomputer," Patent Application Files, 1992.

8.   Roska, T., and Chua, L. O., "The CNN Universal Machine and Supercomputer IEEE Transactions of Circuit and Systems," Part II, March 1993.

9.   Proceedings of 1990 IEEE Second International Workshop on Cellular Neural Networks and their Applications, CNNA '90 Budapest, IEEE Publication No. 90TH0312-9, December 16-19, 1990.

10. Proceedings of 1990 IEEE Second International Workshop on Cellular Neural Networks and their Applications, CNNA '92 Munich, October 14-16, 1992.

11. Special Issue on Cellular Neural Networks, Guest Editor. J. Vandewalle and T. Roska, International Journal of Circuit Theory and Applications, September 1992.

12. Special Issue on Cellular Neural Networks Applications, Guest Editors J. Nossek and T. Roska.

13. Werblin, F., "Synaptic Connections, Receptive Fields and Patterns of Activity in the Tiger Salamander Retina in Investigative Ophthalmology," 32: pp. 459-483, 1990.

14. Roska, T. and Vandewalle, J., Cellular Neural Networks, Editors: John Wiley & Sons, Ltd., London 1993 (to appear).

15. Nasserbakht, G. N., Adkisson, J. W., Wooley, B. A., Harris, J. S., Kamins, T. I., "A Monolithic GaAs-on-Si Receiver Front End for Optical Interconnect System," IEEE Four. of Solid-State Circuits, Vol. 28, No. 6, pp. 622-630, June 1993.

16. Su, D. K., Joinaz, M. J., Masui, S., Wooley, B. A., "Experimental Results and Modeling Techniques for Substrate Noise in Mixed-Signal Integrated Circuits," IEEE Jour. of Solid-State Circuits, Vol. 28, No. 4, pp. 420-430, April 1993.

17. Sevenhans, J., et al., "CMOS LED Driver and PIN Receiver for Fiber Optical Communication at 150 Mbit/sec," Journal of Analog Integrated Circuits and Signal Processing, Vol. 4, No. 1, pp. 31-35, Kluwer Academic Publishers: Boston, MA, July 1993.

18. Muoi, T. V., "Receiver Design for High-Speed Optical-Fiber Systems," IEEE/OSA Jour. Lightwave Technology, Vol. 2, No. 3, pp. 243-267, June 1984.